Homework 4: Q3

Name: Waiwai Kim, waiwaiki

1 Part (a): Proof Idea

We have to prove that $diam(T) = 2d$ for a complete binary tree T. The recitation note uses Lemma 1 and Lemma 2 to prove the following.

$$diam(T) \leq 2D \tag{1}$$

We are tasked with proving the following:

$$diam(T) \geq 2D \tag{2}$$

So that with both equation 1 and 2, we can conclude the following:

$$diam(T) = 2D \tag{3}$$

We will prove Equation 2 by contradiction. Assume that $diam(T) < 2D$. Because we know that we are working with a complete binary tree, we can intuitively say that the longest path is to traverse from the leftmost node to the root of the tree and down to the rightmost node. Because the depth of the tree is d, it follows that either the distance from the root to the leftmost node or the distance from the root to the rightmost node is less than d. This means that at least one node is missing from the leaf. This contradicts our assumption that we are working wit the complete binary tree.

2 Part (b): Proof Idea

We are tasked with take a side whether there exists a constant $c \geq 1$ such that it is the case that

$$\frac{diam(T)}{avgd(T)} \leq c \tag{4}$$

In plain English, we would have to figure out if the maximum distance of any pair in a complete binary tree is always larger than average distance of all possible pairs of two nodes. Is the numerator always bigger than the denominator? First, we know that $diam(T) = 2d$, so we have to express the denominator in terms of d. Secondly, we will define the number of nodes,the number of pairs, the number of pairs with the maximum distance and so on in terms of d or the depth of the tree. Thirdly, we will exploit the fact that we don't have know the exact expression of $avgd(T)$ in order to claim Equation 4 true or false. We can examine if Equation 4 still holds when we take the upper bound of $avgd(T)$ or the largest possible average distance.

3 Part (b): Proof Details

Because we know the numerator, let's focus on the denominator. Denominator can be separated into three different parts: 1) pairs with distance of 1, pairs with distance of $2d$ or the maximum distance and everything else. Equation 5 expresses such. In order to set the upper bound, we could group the sum of everything else into the sum of maximum distance.

$$avgd(T) = \frac{\Sigma_{\{u,v\}\subseteq V}\, dist(u,v)}{\binom{n}{2}} = \frac{\Sigma d = 1 + \Sigma_1^j d = 2d + \Sigma_1^k everything\ else}{\binom{n}{2}} \leq \frac{\Sigma d = 1 + \Sigma_1^{j+k} d = 2d}{\binom{n}{2}} \tag{5}$$

Based on Table 1, we can figure out the number of of elements for each component. If a complete binary tree has depth of d, the tree will have $n = 2^d - 1$ nodes total. There will be total $\frac{n(n-1)}{2}$ pairs, among which $n-1$ pairs have distance of 1. This means our upper bound, we can say that $\frac{n(n-1)}{2} - (n-1)$ pairs have maximum distance of d.

Table 1: Complete Binary tree

d	n = # of nodes $2^d - 1$	# of pairs $\frac{n(n-1)}{2}$	max d $2d$	# of nodes at the bottom $\frac{n+1}{2}$	# of pairs w/ Max D $(\frac{n+1}{4})^2$	# of pairs w/ 1 D $n-1$
1	1	0	0	1	0.25	0
2	3	3	2	2	1	2
3	7	21	4	4	4	6
4	15	105	6	8	16	14
5	31	465	8	16	64	30

Equation 6, we can reduce $avgd(T)$ to a simple form.

$$\frac{\Sigma d = 1 + \Sigma_1^{j+k} d = 2d}{\binom{n}{2}} = \frac{(n-1) + d \cdot (\frac{n(n-1)}{2} - (n-1))}{\binom{n}{2}} = \frac{2}{n} \cdot (1 + \frac{d}{2}(n-2)) = \frac{2 + d(n-2)}{n} \tag{6}$$

Now let's plug the upper bound of $avgd(T)$ to Equation 4. We get Equation 7. The term of $2(d+1)$ become almost irrelevant become the exponential term of 2^d dominates. Let's simply compare $2d(2^d - 1)$ to $d(2^d - 1)$. Because of the coefficient 2, we can conclude that the claim is true that a constant C ≥ 1 exists.

$$\frac{diam(T)}{avgd(T)} \leq \frac{2d}{\frac{2+d(n-2)}{n}} = \frac{2d(2^d - 1)}{2(d+1) + d(2^d - 1)} \tag{7}$$